2018

Monitoring of DNA Polymerase Theta Movements Through FRET Analysis

Ashley Rebelo
ashleymarebelo@gmail.com

Monitoring of DNA Polymerase Theta Movements Through FRET Analysis

By Ashley Rebelo

An Honors Project Submitted in Partial Fulfillment of the Requirements for Honors

In

The Department of Physical Sciences

Faculty of Arts and Sciences

Rhode Island College

2018

Monitoring of DNA Polymerase Theta Movements Through FRET Analysis

An Undergraduate Honors Project Presented

By

Ashley M. Rebelo

To

Department of Physical Sciences

Approved:

_____	_____
Project Advisor	Date
_____	_____
Honors Committee Member	Date
_____	_____
Honors Committee Member	Date
_____	_____
Honors Committee Member	Date
_____	_____
Department Chair	Date

Abstract

DNA polymerases are enzymes used for DNA replication during cell division and can be specialized for DNA repair. DNA Polymerase Theta (Pol θ) is the predominant polymerase involved in alternative double-stranded break repair and is upregulated in breast cancer. It is error-prone as it does not accurately match the nucleotide on a DNA template with the correct complementary base. This inaccuracy affects the overall fidelity of the enzyme, a biochemical process that looks at the ability of a polymerase to "read" the template DNA and select the right nucleotide before polymerization. Polymerization for most DNA polymerases involves a global conformational change, specifically of the fingers domain, during insertion of a nucleotide after selecting it into the DNA strand in the active site. If this selection step is compromised, it could lead to the insertion of a non-complementary nucleotide, which could lead to mutations. The fingers domain is hypothesized to move closer to the active site during this process before release of the extended product, where the fingers domain returns to its original conformation. These results have been previously observed in high-fidelity polymerases β and Klenow. Pol β is involved in BER, and the Klenow fragment synthesizes DNA in *E.coli*. In both studies, the enzyme adopts a closed conformation with correct nucleotide. This study aims to elucidate the mechanism of Pol θ during DNA repair, and how the global movements of this enzyme affects fidelity. Using Fluorescence Resonance Energy Transfer (FRET), we internally fluorescently labeled the fingers domain of Pol θ to observe its interaction with DNA and an incoming nucleotide. Preliminary results suggest that only in the presence of correct nucleotide does Pol θ experience a global conformational change during polymerization to ensure correct matching, suggesting nucleotide selection by Pol θ is monitored in an open conformation and only when the correct pair is formed does polymerization proceed. Our FRET system can be used to further understand the fidelity of Pol θ and how it can lead to cancer.

Table of Contents

Introduction **1**

Experimental Methods **9**

Results and Discussion **15**

Conclusion **24**

Appendix **26**

References **29**

Introduction

Targeting enzymes involved in DNA repair has become an important component in designing cancer therapies. DNA is continuously exposed to environmental factors which can lead to compromising its integrity.

DNA polymerases are an integral part of DNA repair. As their name suggests, these enzymes polymerize DNA, that is they have the ability to synthesis new DNA, either during DNA replication or DNA repair. There are 15 DNA polymerases found in eukaryotes, classified into five families: A, B, X, Y and reverse transcriptase[1]. The variety of these polymerases is due to the specific functions they each have within the cell. These specialized functions are necessary to maintain genomic integrity. A and B family polymerases resemble DNA Polymerase I and II, respectively[2]. Polymerases in the X family do not have homologous sequences with Pol I or II, and Y family polymerases are found in all kingdoms[2]. The Klenow fragment in *E.coli*, which is the large fragment of Pol I, is an A family polymerase that has been well characterized[2], is used for DNA replication. Pol θ, another A-family polymerase[3], is the main enzyme involved in a specific double strand break (DSB) repair pathway called microhomology end-joining (MMEJ)[3]. B-family polymerases such as DNA polymerases are involved in DNA replication in the S phase of the cell cycle[1,4]. DNA polymerase β, an X-family polymerase, is involved in DNA repair, specifically base-excision repair (BER)[5]. Y-family polymerases are involved in translesion synthesis[1]: Pol η can insert the correct A-A substrate opposite the mutant T-T photodimer, also known as thymine dimers[1]. All these DNA polymerases have a right-hand shape and share common domains: fingers (nucleotide binding) domain, a thumb (DNA binding) domain and a palm domain where the active site is located[5]. The fingers domain not only interacts with the incoming nucleotide, but it is also responsible for adding it to the template strand in order to form

the phosphodiester bond during polymerization[6]. The thumb domain ensures the DNA template is correctly aligned with the active site. It is also involved in the processivity of the polymerase during polymerization, which increases as the DNA synthesis increases. Finally, the palm domain catalyzes the formation of the phosphodiester bond between the incoming nucleotide and the 3' end of the DNA strand[6]. Studies on the Klenow fragment and on Pol β[7] have shown that during the polymerization process, DNA first binds to the thumb domain of the polymerase to form a binary complex in an open conformation. The fingers domain selects the correct nucleotide (dNTP), forming a ternary complex[8]. The fingers domain adopts conformational changes (from open to closed) to align the dNTP to the primer DNA within the active site[8], inserts the correct dNTP, allowing the enzyme to catalyze the phosphodiester bond between the 3'OH on the deoxyribose of the primer and the α-phosphate on the 5' end of the incoming dNTP. To stabilize this reaction, two Mg^{2+} metal ions are used: one coordinates the α- phosphate of the dNTP and the 3'-OH of the primer, while the other coordinates the β and γ-phosphate oxygens of the dNTP[8]. The polymerase finally releases the extended DNA product and PP_i by returning to an open conformation before restarting the process (figure 1).

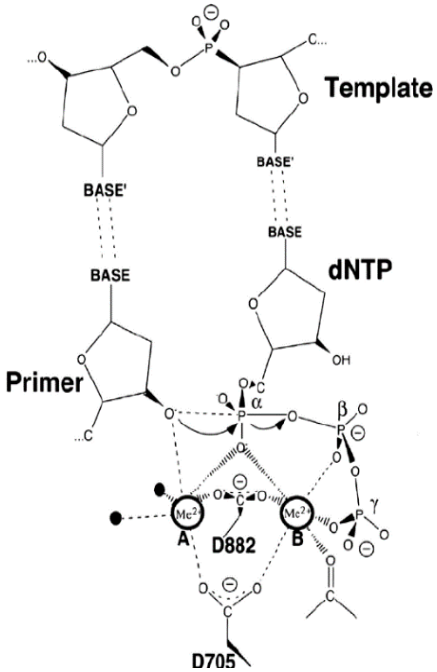


Figure 1. Polymerization Mechanism of DNA Polymerase I. The active site is stabilized with two metal Mg^{2+} ions in octahedral coordination. Each metal ion interacts with amino acid residues of the active site and phosphate groups of the incoming nucleotide[5].

During DNA replication, it is important that the new DNA strand is an exact copy of the original template strand, or mutations can arise from this. This is also true for DNA repair: if a lesion is left unrepaired or misrepaired, it could have

negative consequences on the cell. To overcome this, DNA polymerases such as Pol I have an exonuclease domain (Klenow fragment in Pol I) that are used to proofread the incorporated nucleotide[5]. This results in an error rate of 1 base in ~10^5 bases copied for this enzyme during replication[9], a similar rate that is found in the X-family DNA polymerase β[10]. Interestingly, DNA polymerase θ has an error rate of 1 base in ~400 bases[11] in the MMEJ pathway it is involved in.

The MMEJ pathway can lead to small deletions within the DNA sequence, making it error-prone. It is used as a back-up mechanism when homologous recombination (HR) and classical non-homologous end joining (C-NHEJ) pathways are not used by the cell[6]. HR is considered error-free: this is because HR uses a template strand from an identical sister chromatid[7]. It can therefore only be used when the sister chromatid is present in the cell, which is during the S/G2 phase[7]. HR repair begins with the generation of 3' single-stranded DNA (ssDNA) where the DSB is located[11]. Replication Protein A (RPA) complex then binds to the newly ssDNA, removing any potential secondary structures. BRCA1[7,11] proteins mediates the recruit of Rad51, a DNA strand-exchange protein, which promotes strand invasion into another homologous complex, which is typically the sister chromatid[11]. RPA plays an essential role in this process: it protects and stabilizes the ssDNA, and the HR process cannot continue if this complex is not bond to the ssDNA[11]. The homologous complex is used as a template for DNA synthesis. Once this process is completed, the repaired DNA strand will dissociate and reanneal to the original strand through ligation[7]. If the cell is unable to use this repair pathway, as is the case when the cell is not in the S/G2 phase, it can use alternate repair pathways such as C-NHEJ, which is the fastest way to repair DSB[6]. In this pathway, specialized proteins such as Ku70/Ku80 bind to the DSB site and recruit complexes, such as the DNA-kinase Artemis complex, that remove end groups on each side of the DSB[12]. A DNA polymerase will then synthesize new nucleotides on each end before ligation occurs[13]. If this

pathway is altered, as is the case in Ku-deficient cells, an alternative end joining pathway such as MMEJ is preferred[6]. MMEJ is involved in chromosome rearrangements, and occurs during all phases of the cell cycle[12,13], and is involved in the repair of DNA at collapsed forks during the S phase[8,12]. Studies suggest that it is characteristic in cancer cells[12], indicating that this pathway is used to avoid apoptosis in these types of cells. The MMEJ repair starts with 5'-3' end resection, which exposes microhomologies on the end of each strand[6]. This process is catalyzed in mammalian cells by the MRN complex, Ctp1/Ctp and Mre[8,9], which remove bound proteins and hairpin secondary structures[12]. Once the microhomologies on each end are exposed, end-joining occurs, resulting in the annealing of 5-25 nucleotides in each strand[8,12]. This is catalyzed by Rad52 in vitro for complementary ssDNA higher than 14 nucleotides (nt)[12]. This is followed by removal of non-homologous tails at the 3' end of the newly annealed strands by XPF/ERCC1 in mammalian cells[6]. This allows DNA polymerase θ to fill in the gap before ligation occurs (via DNA ligase III/I)[13]. Pol θ can also catalyze the end-joining step by annealing the microhomologies in each single stranded DNA (ssDNA) before extension of each of these strands *in vitro*[3].

DNA Polymerase Theta (POLQ, Pol θ) is a 2590 amino acid protein[14] located in the nuclei, is categorized as an A-family polymerase, resembles the Klenow fragment[3], and studies have found in green algae *Drosophila*, *C. elegans* and humans that Pol θ is the main enzyme used for the MMEJ repair pathway[3]. Pol θ can do base excision repair (BER) *in vitro* and *in vivo* in *C. elegans,* due to its ability to remove the deoxyribose phosphate (dRP) group in BER[3], suggesting it could be used as a backup for DNA polymerase β[3]. It is also involved in translesion repair, like Y-family polymerases[1], and has the ability to insert adenine residues opposite abasic sites, a result of DNA damage[3]. This upregulation in breast cancer cells have been correlated to a lower survival probability following this diagnosis[3]. It has also been found to contain a mitochondria-targeting

sequence[15], suggesting its involvement in maintaining mtDNA. In fact, this study suggests it has a role when the mitochondria are exposed to oxidative stress[15], suggesting Pol θ is involved in cellular tolerance to this environmental factor. It contains three domains: a N-terminal helicase-like domain, a central domain, and a C-terminal polymerase domain[3]. The helicase-like domain is an ATP-binding domain that is involved in interstrand crosslinks (ICL) and alternative end joining. Studies do not suggest that this domain has helicase activity, but that it has the ability to bind to ssDNA during MMEJ and remove proteins such as Rad51 at the DSB site[16]. The central domain has been hypothesized to play a role in binding proteins involved in the MMEJ pathway[17]. Like other polymerases, Pol θ's polymerase domain (86.6 kDa) is divided into four sub-domains: fingers (nucleotide binding domain), palm (active site domain), thumb (DNA binding domain) and a non-functional exonuclease[3] (figure 2).

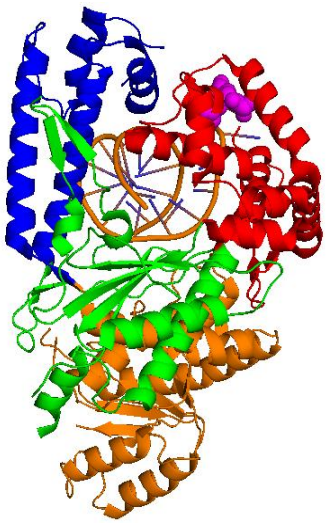


Figure 2. X-ray crystal structure of polymerase domain of Pol θ (4X0Q.pdb)[14]. Structure shows dNTP binding domain on red, DNA binding domain in blue, active site in green and an exonuclease domain in orange. Double stranded DNA is showed as an orange line. The generated Q-tag sequence for FRET labeling is shown in magenta spheres in the fingers domain.

Pol θ is a low fidelity repair enzyme for single base substitutions in vitro[3], which is much higher than its homolog Pol β for BER. Little is known about the fidelity of Pol θ and its nucleotide selection process during DNA repair. Studies also suggest that overexpression of Pol θ in normal cells can lead to more DNA damages, which could indicate that Pol θ is altering genomic integrity[3]. Therefore, the ability to use Pol θ can be an advantage for cells by avoiding apoptosis.

The fidelity of a polymerase describes its ability to go through the previously described mechanism correctly, selecting and inserting the correct nucleotide opposite the template strand via an "induced-fit" mechanism[5]. In normal cells, several factors contribute to the overall accuracy of a DNA polymerase to maintain genomic stability including: 1) free energy difference between complementary and non-complementary base pair, 2) nucleotide discrimination during polymerization within the active site, and 3) proofreading ability of the enzyme, which involves the excision of a potential nucleotide misincorporation and mismatch correction (insertion of the correct nucleotide)[18]. These factors thus minimize the probability of the enzyme to spontaneously mutate DNA during replication or repair, as well as checkpoints during the cell cycle, and therefore making this event rare and random[18]. If the fidelity of an enzyme is compromised, the repair function of the enzyme becomes compromised. In addition, studies done on DNA polymerases suggest that disabling the proofreading ability of the enzyme can compromise its overall accuracy during polymerization, which could in turn increase its mutation rate[18]. This could potentially lead to mutations, which can then accumulate, leading to apoptosis or uncontrolled cellular growth. Enzymatic fidelity of other polymerases, such as the Klenow fragment and Pol β, has been studied using Fluorescence Resonance Energy Transfer (FRET)[9,10]. Both studies have found that nucleotide selection occurs before the first conformational change (from open to closed) occurs, and consequently before nucleotide insertion within the active site of the enzyme[9,10]. This biophysical analysis allows us to observe real-time fluorescent change during nucleotide incorporation in step 3 (fast fingers closing) and 4 (non-covalent step) (figure 3). This method can be used to gain insight into DNA polymerase θ's fidelity, and how cancer cells may use this polymerase to avoid apoptosis. Understanding this polymerase's mechanism of choice would thus

allow us to gain insight into the nucleotide selection process during DNA repair and maintaining genomic stability.

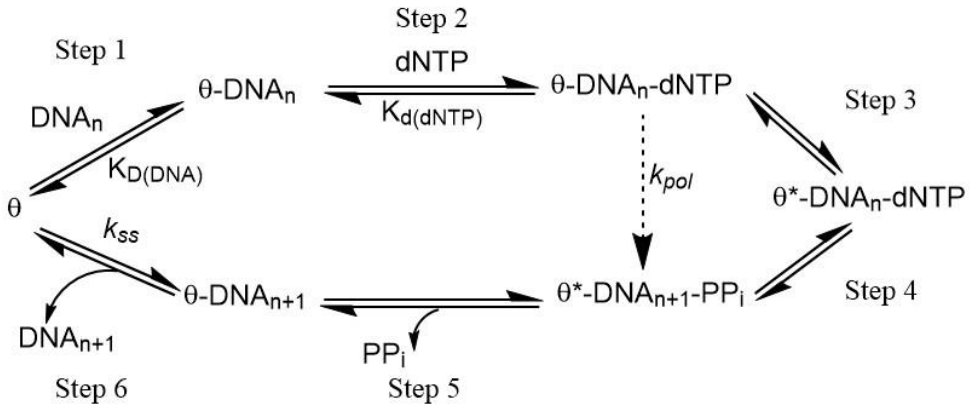


Figure 3. Biochemical mechanism of DNA Polymerase Theta (adapted from Towle-Weicksel et al (2014)[10])

The goal of this project was to study the global movements of Pol θ to gain insight into nucleotide choice and these movements affect fidelity. While some characteristics of this enzyme have been studied, like its crystal structure[14] and its implication in MMEJ[3], little is known about the specifics about its nucleotide selection process. Previous studies on other polymerases by Joyce et al (2008) and Towle-Weicksel et al (2014) showed that during nucleotide selection, the fingers domain remains open in the presence of incorrect nucleotide, but adopts a closed conformation when inserting the correct nucleotide. Both studies suggest that the selection process during polymerization occurs prior to inserting the nucleotide, while the enzyme is still in open conformation[9,10]. To monitor the global movements of these enzymes, a Fluorescent Resonance Energy Transfer (FRET) system was used in those studies to observe a change in fluorescence during nucleotide incorporation[9,10]. This is a non-radiative transfer of energy, in which an excited donor dye (located in the fingers domain of the enzyme) transfers energy to the acceptor dye (located on the DNA). FRET can be used as a molecular ruler to determine the distance (Förster radius) between the donor and the acceptor molecule, thus determining the distance between the

fingers domain of the polymerase and the DNA within the active site. This system can also be used to observe real time fluorescent change when nucleotides are being incorporated, which corresponds to steps 3 and 4 in our hypothesized biochemical pathway of Pol θ (figure 1): step 3 corresponds to the fingers closing after ternary complex (Pol θ-DNA-dNTP) forms, step 4 corresponds to a non-covalent step occurring before release of pyrophosphate, as was found in Pol β (figure 3)[10]. FRET studies on the Klenow fragment and Pol β both labeled the fingers domain with 5-(((2-iodoacetyl)amino)ethyl)aminonaphthalene-1-sulfonic acid) (Iaedans) as the donor dye[9,10], and both studies challenged the induced-fit model. This dye can specifically be added to cysteine residues in a protein. However, we were not able to use this strategy to specifically label the fingers domain of Pol θ because polymerase Theta contains 17 cysteine residues, and mutating 16 cysteines could have compromised the overall activity of Pol θ. Another FRET compatible FRET pair was used: 5-FAM dye was added to the fingers domain of Pol θ, which could then interact with a Dabcyl dye quencher located within a known DNA strand. In order to specifically target the fingers domain of Pol θ for 5-FAM labeling, a specific labeling site (GQQQLG) was generated for a labeling enzyme (transglutaminase) to recognize this sequence and catalyze the addition of the 5-FAM dye to this target, thus specifically labeling the fingers domain where the largest conformational change occurs. By studying the global movements of this enzyme through FRET, we aim to understand how fidelity becomes affected by those movements in the fingers domain. We hypothesize that Pol θ goes through open-closed-open conformational changes, as is observed in other enzymes (Klenow fragment, Pol β). In addition, the ability of Pol θ to do translesion synthesis could suggest its mechanism could be similar to Y-family polymerases also involved in this kind of DNA repair, which includes bypass of thymine dimer and

apurinic/apyrimidinic sites. This could provide significant insight into the nucleotide selection mechanism of Pol θ, and how cancer cells could use this enzyme to avoid apoptosis.

Experimental Methods

Chemical reagents and transglutaminase from guinea pig liver for site-specific labeling were purchased commercially from Sigma Aldrich and AmericanBios. Dabcyl DNA, dideoxy-terminated DNA, deoxynucleotides and dideoxynucleotides used for assays were purchased commercially from IDT (Integrated DNA Technologies). Polyacrylamide gels were scanned using LI-COR Odyssey CLx. All affinity chromatography purifications were done using the AKTA Start Chromatography System (GE HealthCare). pSUMO-hPolQM1 plasmid (8112 bp) was provided by Sylvie Doublié from the University of Vermont. Primers used for site-directed mutagenesis were purchased from Sigma Aldrich and site-directed mutagenesis kit was purchased from Agilent. Mutation of the plasmid for Q-tag Pol θ expression was done by Genewiz.

Generating A2395Q and A2395P

Using 4x0q.pdb crystal structure of Pol θ in PyMol, a potential labeling site was found in the fingers domain, which had the following sequence:

Amino acid #	2394	2399
Amino acid sequence	GAKSLG	

Since the sequence was similar to the 3Q-tag (GQQQLG) used by Lin (2006)[18] we determined that only amino acids AKS (2395 – 2397) needed to be mutated to generate the labeling sequence. The first site directed mutagenesis protocol was done using Agilent QuikChange® II kit (see 1) in appendix).

pSUMO-hPolQM1 plasmid (5ng/uL) was used as a template in polymerase chain reaction, which contained 62.5ng/uL of forward and reverse primers (see appendix for sequences), 1x of QuikChange® II buffer, 100uM of dNTP and 1.25 U of *PfuUltra* High-Fidelity DNA polymerase. PCR sample was heated for 30 seconds at 95°C, then 17 PCR cycles were done with the following steps: 30s at 95°C, 1 min at 55°C, and 12 min at 68°C. The sample was maintained at 4°C until samples were removed from the BioRad PCR instrument. Amplification of the entire sequence was analyzed on a 1% agarose gel with ethidium bromide, under UV light on Bio Rad ChemiDoc XRS+ system (figure 4). The same method was used to generate A2395P (see primers in 2) in appendix) using the same plasmid (figure 4).

Q-tag Pol θ expression

Genewiz generated the modified plasmid with the Q-tag sequence (GQQQLG) engineered in the fingers domain of Pol θ, the plasmid was transformed into Rosetta2(DE3) *E. coli* cells (Novagen) (30s at 42°C in water bath, on ice for 2 minutes) and incubated in LB broth for 1 hour in a 37°C shaker at 225rpm. Cells were then plated on LB-agar lennox plates with Ampicillin (AMP, 100ug/mL) and incubated overnight at 37°C. Colonies were inoculated in 1L Terrific Broth (AmericanBios) with AMP (100ug/mL) and chloramphenicol (CM, 34ug/mL). Inoculated bacterial cultures were incubated at 20°C and grown for 68 hours at 200rpm. Bacterial cultures were centrifuged at 4,000g using the Sorvall Superspeed RC-2 centrifuge, at 4°C. Resulting bacterial pellets were stored in -80°C freezer.

Q-tag Pol θ purification

Purification and labeling processes were done at 4°C. Q-tag Pol θ pellet was thawed overnight at 4°C the day prior to purification. Once the pellet was thawed, the cells were resuspended in lysis

buffer pH 7 (20mM Tris pH 7, 300mM NaCl, 0.01% (v/v) NP-40, 10% (v/v) glycerol, 120uL of 100mM of PMSF, 5mM BME, 1x PIC). The resuspended pellet was sonicated 6 times for 30s (amplitude 21%, 0.5s pulse ON, 1.0s pulse OFF) using a Branson Digital Sonifier 250 to lyse the cells open. The lysed pellet was centrifuged twice for 30 minutes using the Sorvall RC-5C Plus centrifuge (SLA-600TC rotor, 4°C, 16,100g). The supernatant was separated on a nickel-charged HisTrap FF Crude affinity column on an AKTA Start Chromatography System (GE HealthCare), using binding buffer F pH 7 (20mM Tris pH 7, 300mM NaCl, 0.01% (v/v) NP-40, 10% (v/v) glycerol, 20mM imidazole, 5mM BME) and elution buffer B pH 7 (20mM Tris pH 7, 300mM NaCl, 0.01% (v/v) NP-40, 10% (v/v) glycerol, 500mM imidazole, 5mM BME). Pooled fractions from the corresponding peak were run on a HiTrap Heparin affinity column to remove non-DNA binding protein, using binding buffer C pH 7 (20mM Tris pH 7, 300mM NaCl, 0.01% (v/v) NP-40, 10% (v/v) glycerol, 5mM BME) and elution buffer E pH 7 (20mM Tris pH 7, 2M NaCl, 0.01% (v/v) NP-40, 10% (v/v) glycerol, 5mM BME) using the same chromatography instrument.

Q-tag Pol θ labeling and purification

Pooled fractions from the corresponding peak of the Heparin purification were reduced using 0.01M DTT and incubated at 4°C for 20 minutes. Reduced protein was incubated at 4°C for 2 hours in labeling buffer (0.1uM 5-FAM cadaverine, 0.004uM transglutaminase, 10mM $CaCl_2$, 20mM Tris pH 7) (see labeling schematic 5) in appendix). The labeling reaction was quenched with 1M DTT and incubated at 4°C for 30 minutes. Labeled Q-tag was then run on a HiTrap Chelating affinity column to separate the labeled protein from the transglutaminase and excess 5-FAM cadaverine. The protein solution was bound to the column using buffer A pH 7 (20mM Tris pH 7, 300mM NaCl, 0.01% (v/v) NP-40, 10% (v/v) glycerol, 10mM imidazole, 5mM BME) and

eluted in buffer B. Fractions from the corresponding peak were pooled together and Ulp1 SUMO protease (10:1 mass ratio) was added for overnight cleavage at 4°C of the 6xHis-SUMO tag.

Cleaved labeled protein was run on a nickel-charged HiTrap Chelating affinity column to separate the labeled protein from the 6xHis-SUMO tag, using binding buffer A and elution buffer B. Fractions from the corresponding peak were pooled together, flash frozen in liquid nitrogen and stored in -80°C freezer.

WT Pol θ was purified and labeled using the same protocol as for Q-tag Pol θ to run as a negative control. The presence of labeled protein was confirmed on an SDS-PAGE gel using the Molecular Dynamics Typhoon 9410 scanner for fluorescence visualization, and Coomassie stained for protein visualization (figure 6). Loaded samples contained protein mixed with loading buffer (65.8mM Tris pH6.8, 26.3% (v/v) glycerol, 2.1% SDS, 0.71M beta-mercaptoethanol (BME)). Labeling efficiency was Q-tag Pol θ was determined using a coupling efficiency (CE, or labeling efficiency) equation[19], where CE is equal to:

(Equation 1)
$$\frac{\varepsilon_{280}(\text{Pol }\theta) \cdot A_{5FAM}}{(A_{280} - CF_{280} \cdot A_{5FAM}) \cdot \varepsilon_{5FAM}}$$

CF_{280} is defined as the correction factor, $CF_{280} = A_{280}$ (free 5FAM dye) /A_{5FAM} (free 5FAM dye) = 0.33, ε_{280}(Pol θ) = 68,300 M^{-1} cm^{-1}, ε_{5FAM} = 60,000 M^{-1} cm^{-1}.

Pre-steady state kinetics

5’-IR(800nm) tagged single strand DNA primer was annealed to a complementary extendable DNA template strand, of 17 bp and 20 bp respectively (see 4) in appendix). 1uM primer, 1.2uM template and annealing buffer (50mM Tris pH 6.8, 250mM NaCl) were incubated at 95°C for 5

minutes. DNA samples were then gradually cooled on the benchtop for overnight annealing. The duplexed DNA was analyzed and confirmed using 12% Native Page gel on a LI-COR Odyssey CLx laser scanner in the 700-nm channel (data not shown).

100 nM of Pol θ was pre-incubated with 250 nM duplex DNA with running buffer. This was mixed with 500 uM dATP in 10mM $MgCl_2$, running buffer and formamide dye using RQF-3 (KinTek Corp) at 37°C over various time points going from 0s to 10s. Reactions were quenched with 0.26M EDTA pH8 and products were separated on a 20% denaturing polyacrylamide gel. Data was quantified using ImageStudio (LI-COR) and plotted in Prism GraphPad. Time points of product formation were fit to the following non-linear regression equation to determine the amplitude of product release (or burst) rate of polymerization (k_{obs}), product release (k_{ss}) and the amount of active enzyme (E_{app})[10].

(Equation 2) $$[Product] = [E]_{app}\left[\frac{{k_{obs}}^2}{(k_{obs}+k_{ss})^2} \times \left(1 - e^{-(k_{obs}+k_{ss})t}\right) + \frac{(k_{obs}k_{ss})}{(k_{obs}+k_{ss})}t\right]$$

FRET analysis

A single strand ddDNA (dideoxy DNA) primer was annealed to a complementary non-extendable DNA template strand, of 26 bp and 40 bp respectively (see 4) in appendix), using the same procedure as for the annealed DNA used for pre-steady state kinetics. The duplexed DNA was analyzed using a 12% Native Page gel in ethidium bromide and confirmed under UV light on Bio Rad ChemiDoc XRS+ system (data not shown).

Solutions prepared for this assay all contained 50mM Tris pH 7.6, 10mM $MgCl_2$, 1mM EDTA pH 8 and 30nM 5-FAM Q-tag Pol θ. The following was added to each set of tubes: 0.1uM DNA, then either 100uM incorrect nucleotide (dATP, dCTP) or 100uM correct nucleotide (dGTP). The

fluorescent emission of these samples was collected using a PTI Felix 32 Spectrofluorimeter at $\lambda_{excitation}$ = 470 nm at Yale School of Medicine and plotted in Prism GraphPad.

Results and Discussion

Generating Q-tag for fluorescence labeling of Pol θ

The fingers domain of Pol θ is located from Q2333 to Q2474 in the amino acid sequence of this protein (4X0Q.pdb[14]). R2379, K2383 and Y2387 are highly conserved residues essential (see appendix), as they bind to the incoming nucleotide and interact with the active site during nucleotide incorporation[14], and could therefore not be modified for fluorescent labeling. The amino acid sequence GAKSLG (G2394 to G2398) was determined to be an optimal labeling site, as it did not involve mutating residues involved in nucleotide binding, but was close enough to this site to observe the biggest FRET change. In addition, generating the triple mutation A2395Q, K2396Q and S2397Q would generate the labeling sequence GQQQLG, which has already been done by Lin *et al* and showed a labeling efficiency of 70%[18]. In order to optimize site directed mutagenesis, we determined that changing one amino acid at a time would result in a higher probability of the mutagenesis to occur during PCR. Mutation of the AKS sequence to QQQ using site-directed mutagenesis proved to be difficult. Results from PCR of A2395Q shows a band below 0.5 kilobases (kb), suggesting this band corresponds to the primers used for PCR amplification (figure 4A). However, there was no band at ~8kb, which is where we expected our amplified mutated plasmid to be.

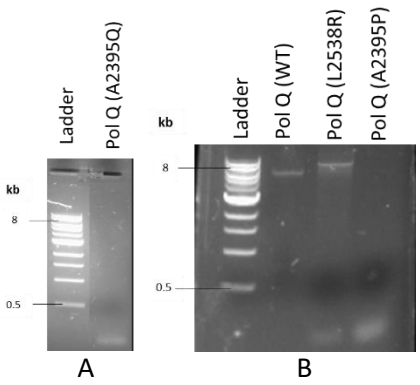


Figure 4. 1% agarose gel of PCR amplification of POLQ mutations. A: POL Q (A2395Q) lane shows a band at ~ 0.1kb. B: POL Q WT lane shows a band at 8 kb, POL Q (L2538R) lane shows two bands (0.1 kb and 8kb) and POL Q (A2395P) lane shows one band at ~ 0.1kb.

Since the resulting PCR was unsuccessful (figure 4A), parameters in the protocol were changed to optimize PCR: the denaturing temperature was changed from 95°C to 98°C for 10s, with an extension temperature of 72°C for 4 minutes instead of 68°C for 12 minutes, with no annealing step (protocol from Invitrogen). The cycle was repeated 30 times using Platinum SuperFi DNA Polymerase (Invitrogen). These changes were unsuccessful (data not shown). An alternative site-directed mutagenesis protocol was used to generate A2395Q using single-primer PCR (Edelheit) using the initial polymerase (PfuTurbo), but was also unsuccessful (data not shown). After these unsuccessful repeats, we hypothesized that the primers used could have been at fault because it involved changing all three nucleotides within the codon (GCT → CAG), which could have affected annealing of the primers to the plasmid DNA. To overcome this, new primers were designed to only change one nucleotide at a time, starting with the first one in the original codon (GCT → CCT) (see 2) in appendix). The initial site-directed mutagenesis protocol was followed (see experimental methods), and this same protocol was done to generate L2538R (figure 4B). Generating this variant has already been done using the same site-directed mutagenesis kit, and was therefore run as a positive control to ensure that the QuikChange® II kit from Agilent was not at fault. Wild-type (WT) POLQ was run as a control, and the band at 8 kb is where we expected it

to be, as was the case for the amplified L2538R sample. The lower band in L2538R at ~0.1kb corresponds to the primers used during PCR. Only a band corresponding to the primers was visible in the A2395P sample. Results from the PCR amplification of L2538R and A2395P suggest that the kit is not at fault, but the amplification and mutagenesis of the first amino acid to generate the Q-tag sequence was still unsuccessful.

Q-tag Pol θ is purified under similar conditions as WT

The Genewiz, LLC lab generated the modified plasmid and confirmed presence of the generated mutation. The following expression process used was adapted to what has already been reported in the literature method[20]: the cell lysate was purified using only two successive affinity chromatographies (nickel and heparin). Separation of Q-tag Pol θ from cleaved 6xHis-SUMO tag was then done using another nickel affinity chromatography (see experimental procedure). Successful purification and cleavage of the 6xHis-SUMO tag were confirmed by SDS-Page (figure 5). Coomassie stained SDS-PAGE shows a band above 100kDa in lane 1 (Q-tag HEP). This was to be expected since this sample was from the pooled elution fractions after heparin affinity chromatography, before the 6xHis-SUMO tag was removed (101.3 kDa). Fractions 3 to 8 from nickel affinity chromatography after 6xHis-SUMO cleavage shows a single expected band between 75 and 100 kDa, which is where Q-tag Pol θ (86.6kDa) was expected to migrate, therefore confirming purification of this protein was successful.

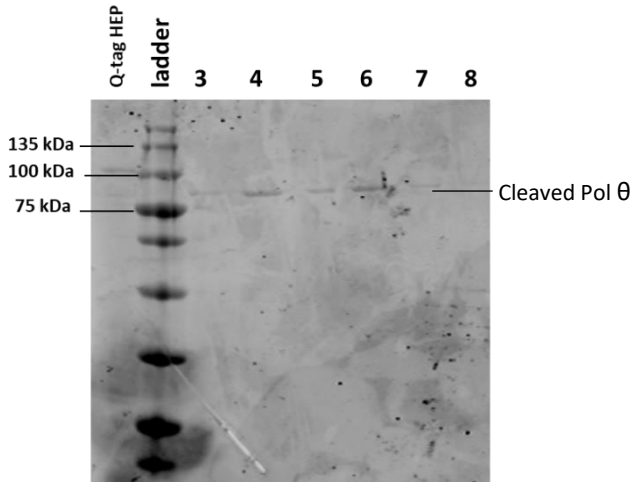


Figure 5. Q-tag Pol θ is highly purified following FPLC purification. Samples from the last purification process of Q-tag Pol θ were separated by SDS-PAGE and visualized via Coomassie stain on a LI-COR scanner. Lane 1: Q-tag Pol θ before 6xHis-SUMO tag cleavage; lanes 3 through 8: elution fractions from the last purification chromatography after 6xHis-SUMO tag cleavage.

Unlabeled Q-tag Pol θ does not show significantly altered activity

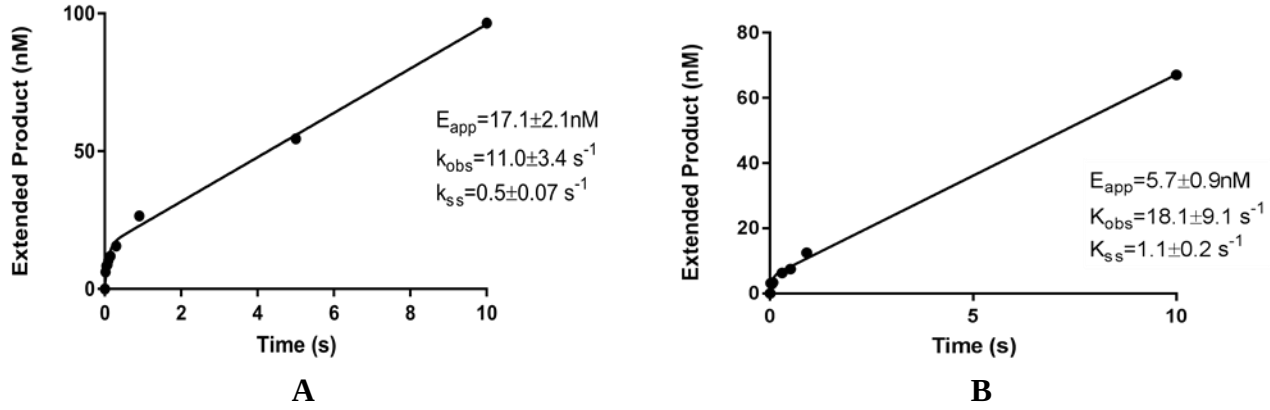


Figure 6. Modified Pol θ displays comparable burst activities as WT Pol θ. A) Pre-steady state kinetics of WT Pol θ. B) Pre-steady state kinetics of unlabeled Q-tag Pol θ. Both pre-steady state assays were performed on a 20mer DNA containing a 5-nucleotide gap, and each time point was fit to equation 2.

Pre-steady state conditions were applied to Q-tag Pol θ to determine if the generated mutation did not alter the overall activity of Q-tag Pol θ (figure 6) compared to WT Pol θ. E_{app}, the amount of active enzyme, of Q-tag Pol θ is lower than that of WT, suggesting that there is less active Q-tag Pol θ. This could be because the purified Q-tag Pol θ used for this pre-steady state burst had not been recently purified and could have started to degrade. Also, the secondary structure of Q-tag Pol θ could have been altered compared to that of WT, resulting in a decrease

in amount of active enzyme. It is difficult to determine if both initial rates of polymerization k_{obs} are similar, due to the margin of error of both WT and Q-tag Pol θ overlapping. The rate of product release k_{ss} suggests Q-tag Pol θ is releasing extended product twice as fast as WT. However, modified Q-tag Pol θ displayed a classic biphasic burst trend, as did WT Pol θ, suggesting that the mutation did not significantly alter the activity of the enzyme. Both WT and modified Q-tag Pol θ display similar biphasic kinetics.

Labeling with 5-FAM of Q-tag Pol θ shows fluorescent signal

To first optimize labeling conditions of Pol θ, a time-course was done on purified samples to determine the optimal time for maximum labeling to occur. The labeling process was done on both Q-tag Pol θ and WT Pol θ as a negative control (figure 7). Two bands were identified (above 75 kDa and at 75kDa) on fluorescently analyzed SDS-PAGE gel in Q-tag samples and one band in WT samples (figure 7B). The higher band is hypothesized to be self-labeling of transglutaminase (76.6 kDa), explaining the presence of an extra band in Q-tag samples. This indicates that time-course fractions from WT do not show evidence of labeling. Fractions from Q-tag show the transglutaminase labeled protein migrating faster than the transglutaminase.

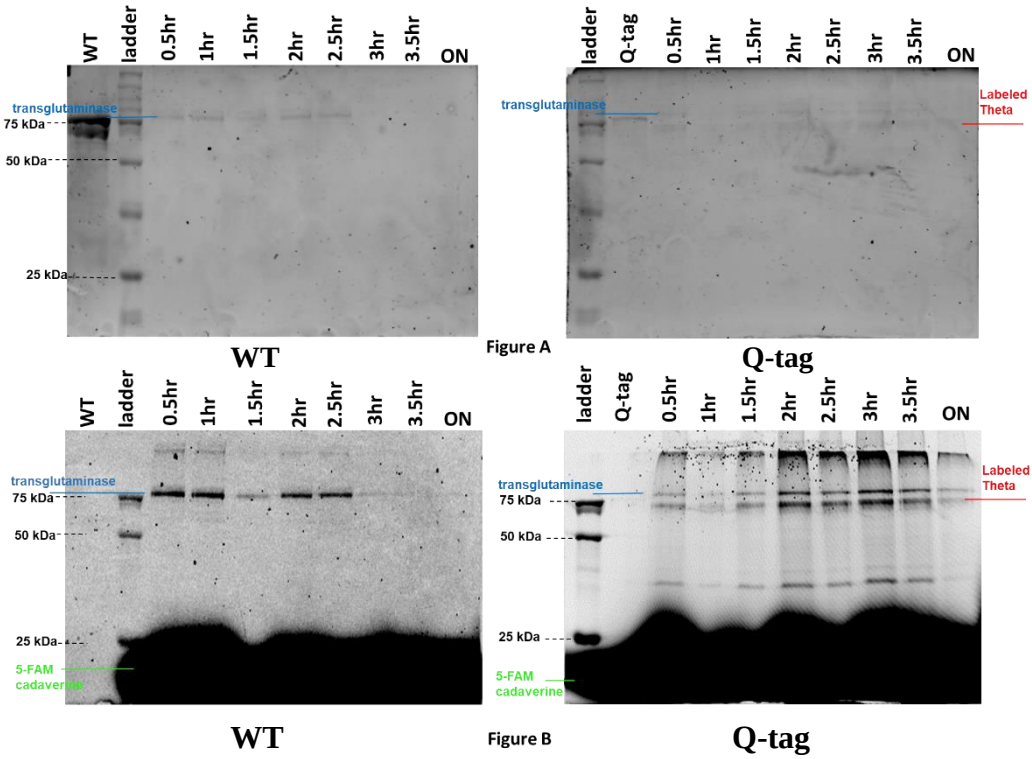


Figure 7. Time-course labeling of Pol θ results in Q-tag labeling. Samples from the time-course reaction of WT and Q-tag Pol θ were separated by SDS-PAGE and visualized via Coomassie stain on a LI-COR scanner (figure A) and fluorescence on a Typhoon 9400 laser scanner (figure B).

Because of the self-labeling ability of transglutaminase and the presence of excess cadaverine (figure 7), it became necessary to remove this enzyme in order not to alter FRET results. Labeling of Q-tag Pol θ was therefore done concurrently with the purification process to remove any excess 5-FAM cadaverine and transglutaminase (see experimental methods). Purification of Q-tag Pol θ was changed from the literature[20] to combine both purification and labeling processes in order to obtain highly purified labeled Pol θ, during which labeling of Q-tag Pol θ occurred before cleaving 6xHis-SUMO tag (see experimental methods). As was done during the time course, WT also went through the same purification and labeling process as the Q-tag. Samples from the labeling process were run separated via SDS-Page and site-specific labeling of the Q-tag was verified (figure 8).

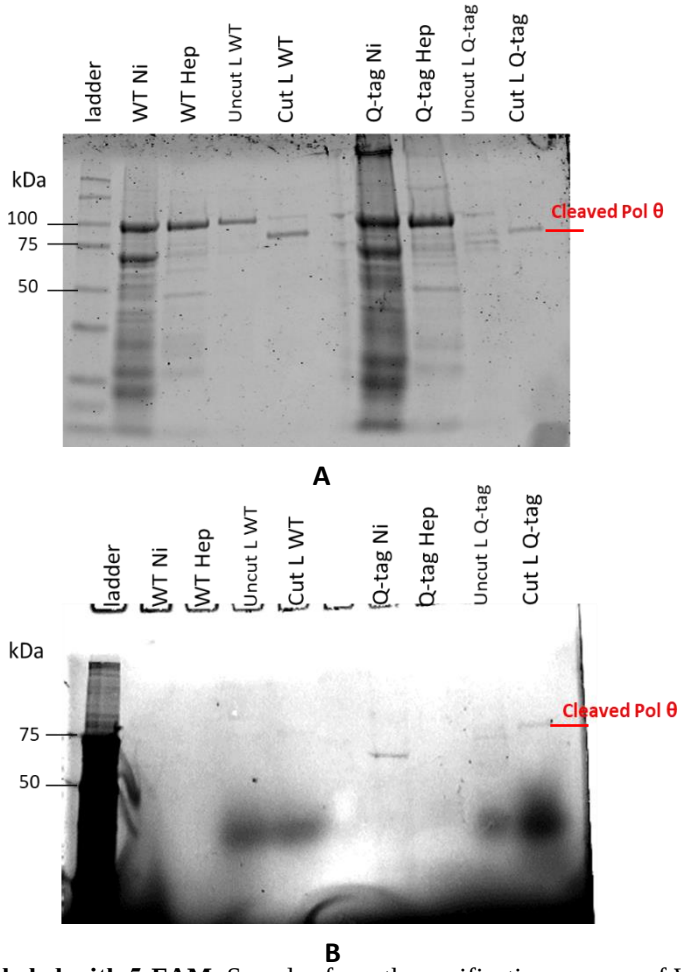


Figure 8. Q-tag Pol θ is labeled with 5-FAM. Samples from the purification process of WT and Q-tag Pol θ were separated by SDS-PAGE and visualized via Coomassie stain on a LI-COR scanner (figure A) and fluorescence on a Molecular Dynamics Typhoon 9410 scanner (figure B). Fractions from uncut L (labeled) WT and cut L WT show no evidence of labeling, whereas cut L Q-tag fraction shows fluorescent protein.

Fluorescent analysis at 470nm of the SDS-Page gel shows a band in the cleaved Q-tag sample. There is a faint band in the uncut labeled WT sample. This could be because the SUMO2 sequence in the 6xHis-SUMO tag contains 3 consecutive glutamines, as is the case for our labeling site[21] (see appendix). Coomassie staining of this same gel shows bands in all samples, confirming the presence of protein in every loaded well. However, the amount of protein loaded was not normalized, therefore the amount of protein shown after Coomassie staining varies in each well. Cleaved Pol θ migrated between 75 and 100 kDa, which is what was expected after cleavage of the 6x-His SUMO tag (see appendix). The visible band in the fluorescent analysis is also above 75 kDa, confirming that Q-tag Pol θ was successfully labeled at the generated site with 5-FAM and

that the transglutaminase was successfully removed from both WT and Q-tag samples. Figure 8B also shows a band in Q-tag Ni sample at ~75kDa, which is not expected since the protein does not go through the labeling process until after heparin purification. This could be due to a loading issue. Since there is no bromophenol blue in the loading buffer of the SDS-PAGE samples, loading samples into each well proved to be difficult to visualize. Therefore, this Q-tag Ni sample could have been mislabeled as another sample that contained labeled protein. Using equation 1[19], labeling efficiency of Q-tag Pol θ was 5.88% after final purification.

Pol θ displays a closed-fingers conformation in the presence of correct nucleotide

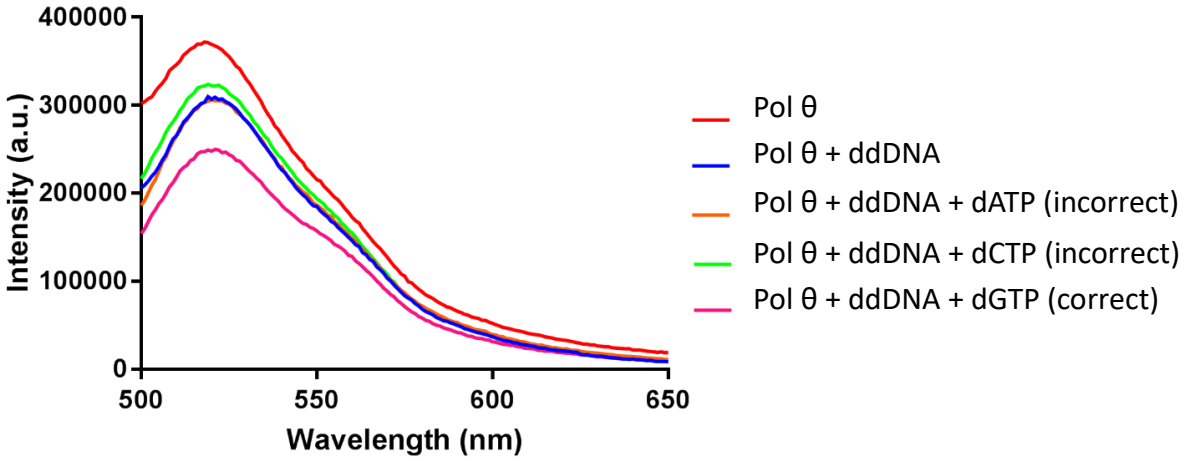


Figure 9. Steady State FRET emission spectra (470 nm) of Pol θ suggest a closed confirmation when incorporating the correct nucleotide. Samples were visualized using a PTI Felix 32 Spectrofluorimeter, at $\lambda_{excitation}$ = 470nm[22]. Fluorescence signal of 600nM Pol θ (red) is reduced when in the presence of 100nM Dabcyl ddDNA (blue). This signal remains the same when incorrect nucleotide (100uM dATP or 100uM dCTP, orange and green, respectively) is added. Fluorescence intensity decreases further when Pol θ and ddDNA are in the presence of correct nucleotide (100uM dGTP, pink).

Following 5-FAM labeling of Q-tag Pol θ, FRET analysis at 470nm[22] was performed using a 40 mer non-extendable Dabcyl-DNA strand with a 20-nucleotide overlap (see 4) in appendix). Previous studies by Joyce et al have determined that presence of Dabcyl quencher 8 nucleotides before the gap within the DNA sequence was optimal FRET analysis[9]. This non-extendable Dabcyl-DNA was used with a deoxy-nucleotide of choice to observe nucleotide incorporation

without any extended product. As the dyes on the protein and quencher on the DNA are closer (within the Förster radius) during nucleotide incorporation, the energy from the excited 5-FAM on Pol θ will get transferred to the Dabcyl dye on the DNA, resulting in a decrease in fluorescent signal (figure 9). As expected, Pol θ emitted at ~518 nm, when no other substrate was present. When non-extendable Dabcyl-DNA was added, the signal decreased at ~518 nm, suggesting that the dye in the fingers domain of Pol θ and the Dabcyl in the DNA are close, and therefore that the enzyme is binding to the DNA, forming a binary complex. When adding incorrect nucleotide (dATP and dCTP) to Pol θ and non-extendable DNA, the fluorescence intensity did not change compared to the binary complex. However, the signal decreased significantly in the presence of correct nucleotide (dGTP), suggesting the dyes are in closer proximity with each other, thus a closed conformation of the overall protein. This data suggests that the enzyme is using different mechanisms when selecting the correct nucleotide as opposed to the incorrect nucleotide, which is different from what was observed in DNA polymerase β[10] but similar to what is observed in the Klenow fragment[9]. This fingers-closing process only in the presence of correct nucleotide suggests that the induced-fit model during nucleotide incorporation in DNA repair is not applicable in the case of Pol θ, as it does not adopt a closed conformation regardless of the nucleotide present.

Conclusion

Pol θ is involved is MMEJ, a specific DSB repair pathway used when homologous recombination and C-NHEJ pathways are compromised. Because Pol θ is a low fidelity enzyme, and has a high error rate during DNA repair, which suggest it could lead to mutations within the cell that could potentially lead to uncontrolled cell growth, as suggested by overexpression of Pol θ in certain types of cancer. Although its involvement in MMEJ has been established, understanding the fidelity of this enzyme and the mechanisms involved during nucleotide selection has not been determined, and could provide insight into this process. We expressed modified WT Pol θ for site-specific labeling of a 5-FAM, which was then used in FRET analysis to gain insight into the nucleotide selection process. We have successfully purified and added a 5-FAM dye in the nucleotide binding domain of Pol θ, and subsequently confirmed the labeling site. Furthermore, we confirmed modification of the amino acid sequence for site-specific labeling did not affect the overall activity of Pol θ. Finally, FRET assays demonstrate that the fingers domain of Pol θ interact with DNA only in the presence of correct nucleotide, suggesting that Pol θ selects the correct nucleotide to insert prior to inserting it in the active site, similarly to the Klenow fragment. This challenges the induced-fit model in which a DNA polymerase always adopts a closed conformation during polymerization when selecting the correct substrate. This suggests that the nucleotide selection process occurs before nucleotide incorporation, not while this process occurs, and that different mechanisms are involved in this selection process. However, using FRET for enzymatic studies is limited because of the nature of the study: if the donor dye is exposed to light over an extended period of time, the results can be altered due to photobleaching of the dye. This becomes problematic since, after some time exposed to light, it is difficult to determine if he signal emitted is due to photobleaching of the donor dye or due to the quencher dye. Using FRET for this study

is also limited because of the labeling process. Although a specific sequence in the fingers domain was generated for the labeling enzyme to specifically catalyze the addition of the donor on it, there is another QQ sequence 15 amino acids before the target: this could have been a potential target for the labeling enzyme as well. We are currently confirming the location of the 5-FAM dye on the generated Q-tag in the fingers domain. In order for a significant FRET change to occur, both dyes need to be within a distance of 10 to 100Å, and confirmation of the location of the 5-FAM dye will allow us to determine the Förster radius between both dyes in the fingers domain and the DNA. The next step will be determining the stopped flow rates of the closing/opening of the fingers domain, and if an additional non-covalent step is part of the polymerization process. Finally, this FRET model will be applied to Pol θ cancer variants to understand how the nucleotide selection mechanism of these variants could contribute to higher mutation rates, and how potential alteration of this mechanism can be a contributing factor to cancer cells.

Appendix

1) Primers used for 1st site-directed mutagenesis: changing one amino acid at a time

A2395Q	F 5'-CCATCTGCTCTCCCAAAGATTT**ctg**TCCCATTCCATAAATGATCCCA-3' R 5'-TGGGATCATTTATGGAATGGGA**cag**AAATCTTTGGGAGAGCAGATGG-3'
K2396Q	F 5'-CCATCTGCTCTCCCAAAGA**ctg**CTGTCCCATTCCATAAATG-3' R 5'-CATTTATGGAATGGGACAG**cag**TCTTTGGGAGAGCAGATGG-3'
S2397Q	F 5'-AATGCCATCTGCTCTCCCAA**ctg**CTGTCCCATTCCATAAAT-3' R 5'-ATTTATGGAATGGGACAGCAG**cag**TTGGGAGAGCAGATGGGCATT-3'

F: forward primer

R: reverse primer

2) Primers used for 2nd site-directed mutagenesis: changing one nucleotide in the codon at a time

A2395Q

A2395P **G**CT→**C**CT	F 5'-TGCTCTCCCAAAGATTT**agg**TCCCATTCCATAAATGATC-3' R 5'- GATCATTTATGGAATGGGA**cct**AAATCTTTGGGAGAGCA-3'
P2395H C**C**T→C**A**T	F 5'- TCTGCTCTCCCAAAGATTT**atg**TCCCATTCCATAAATGATC-3' R 5'- GATCATTTATGGAATGGGA**cat**AAATCTTTGGGAGAGCAGA-3'
H2395Q CA**T**→CA**G**	F 5'- CATCTGCTCTCCCAAAGATTT**ctg**TCCCATTCCATAAATGATC-3' R 5'- GATCATTTATGGAATGGGA**cag**AAATCTTTGGGAGAGCAGATG-3'

F: forward primer

R: reverse primer

3) Amino Acid sequence of 6xHis-SUMO tagged Q-tag Pol θ

MGHHHHHH**GSLQEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMK**
AYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGGFKDNSPIS
DTSFSLQLSQDGLQLTPASSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLT
SSKTATIGSRFKQASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGGRDAYYFSLQKEQKH
SEISASLVPPSLDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQS
YEDPKVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLGLNAGSEHSGRYR

ASVESILIFNSMNQLNSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQKHI
MQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKTLGSTRRGI
DNGRKLRLGRQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVVFPLQREKCLNPFLG
MERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPMGRGKY
KKGFSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSH
DRRLIQVLNTGADVFRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGM**GQQQLG**EQM
GIKENDAACYIDSFKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIKDNNPYRK
AHAERQAINTIVQGSAADIVKIATVNIQKQLETFHSTFKSHGHREGMLQSDRTGLSRKRK
LQGMFCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGA
SWGELKDFDV

Underlined: 6xHistidine tag

In bold: SUMO tag

Underlined in red: SUMO2 protease cleavage site

Yellow highlights: conserved residues in fingers domain

In red: Q-tag location (fingers domain)

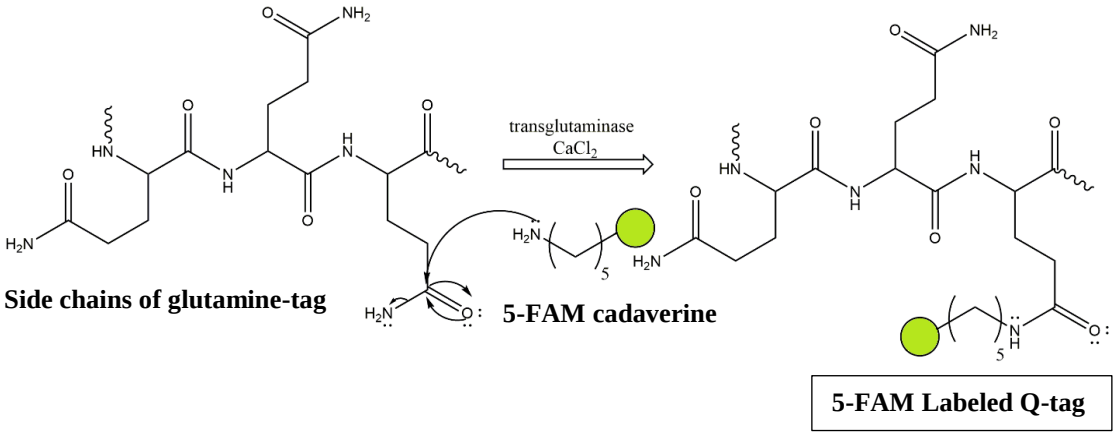
4) DNA substrates

Substrate	**DNA sequence**
T:3'OH[A]	5'-/IRD800/TTTGCGGCTATCATAAG -3' 3'-CGCCGATAGTATTCTACCA -5'
T(-8) D:3'H[B]	5'-TTTGCCTTGCCATGTAACAGAGAGC$_{dd}$ -3' 3'-CGGAACTGGTACAT**X**GTCTCTCGCACTCACTCTCTTCTCT-5'

[A] T:3'OH is extendable DNA substrate without Dabcyl residue

[B] T(-8)D:3'H is non-extendable DNA (3'-dideoxycytidine) with Dabcyl residue (X)

5) Labeling schematic



Side chains of glutamine-tag **5-FAM cadaverine**

5-FAM Labeled Q-tag

References

[1] Cotterill, S., & Kearsey, S. (2002). Eukaryotic DNA Polymerases. Encyclopedia of Life Sciences.

[2] Shanbhag, V., Sachdev, S., Flores, J., Modak, M., & Singh, K. (2018). Family A and B DNA Polymerases in Cancer: Opportunities for Therapeutic Interventions. *Biology*, 7(1), 5.

[3] Beagan, K., & Mcvey, M. (2015). Linking DNA polymerase theta structure and function in health and disease. *Cellular and Molecular Life Sciences*, 73(3), 603-615.

[4] Lujan, S. A., Williams, J. S., & Kunkel, T. A. (2016). DNA Polymerases Divide the Labor of Genome Replication. Trends in Cell Biology, 26(9), 640-654.

[5] Steitz, T. A. (1999). DNA Polymerases: Structural Diversity and Common Mechanisms. Journal of Biological Chemistry, 274(25), 17395-17398.

[6] Wang, H., & Xu, X. (2017). Microhomology-mediated end joining: New players join the team. *Cell & Bioscience*, 7(1).

[7] Prakash, R., Zhang, Y., Feng, W., & Jasin, M. (2015). Homologous Recombination and Human Health: The Roles of BRCA1, BRCA2, and Associated Proteins. *Cold Spring Harbor Perspectives in Biology*, 7(4).

[8] Truong, L. N., Li, Y., Shi, L. Z., Hwang, P. Y., He, J., Wang, H., . . . Wu, X. (2013). Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences*, 110(19), 7720-7725.

[9] Joyce, C. M., Potapova, O., Delucia, A. M., Huang, X., Basu, V. P., & Grindley, N. D. (2008). Fingers-Closing and Other Rapid Conformational Changes in DNA Polymerase I (Klenow Fragment) and Their Role in Nucleotide Selectivity†. *Biochemistry*, 47(23), 6103-6116.

[10] Towle-Weicksel, J. B., Dalal, S., Sohl, C. D., Doublié, S., Anderson, K. S., & Sweasy, J. B. (2014). Fluorescence Resonance Energy Transfer Studies of DNA Polymerase β. *Journal of Biological Chemistry*, 289(23), 16541-16550.

[11] Krejci L, Altmannova V, Spirek M, Zhao X. Homologous recombination and its regulation. *Nucleic Acids Res*. 2012; 40(13):5795–5818.

[12] Sfeir, A., & Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences*, 40(11), 701-714.

[13] Kelley, M. R., Logsdon, D., & Fishel, M. L. (2014). Targeting DNA repair pathways for cancer treatment: What's new? *Future Oncology*, 10(7), 1215-1237.

[14] Zahn, K. E., Averill, A. M., Aller, P., Wood, R. D., & Doublié, S. (2015). Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. *Nature Structural & Molecular Biology*, 22(4), 304-311.

[15] Newman, J. A., Cooper, C. D., Aitkenhead, H., & Gileadi, O. (2015). Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway. *Structure*, 23(12), 2319-2330.

[16] Malaby, A. W., Martin, S. K., Wood, R. D., & Doublié, S. (2017). Expression and Structural Analyses of Human DNA Polymerase θ (POLQ). *Methods in Enzymology DNA Repair Enzymes: Structure, Biophysics, and Mechanism*, 103-121.

[17] Loeb, L. A., Loeb, K. R. & Anderson, J. P. (2003). Multiple mutations and cancer. *Proceedings of the National Academy of Sciences*, 100(3), 776-781.

[18] Lin, C., & Ting, A. Y. (2006). Transglutaminase-Catalyzed Site-Specific Conjugation of Small-Molecule Probes to Proteins in Vitro and on the Surface of Living Cells. *Journal of the American Chemical Society*, 128(14), 4542-4543.

[19] Kim, Y., Ho, S. O., Gassman, N. R., Korlann, Y., Landorf, E. V., Collart, F. R., & Weiss, S. (2008). Efficient Site-Specific Labeling of Proteins via Cysteines. *Bioconjugate Chemistry*, 19(3), 786-791.

[20] Hogg, M., Seki, M., Wood, R. D., Doublié, S., & Wallace, S. S. (2011). Lesion Bypass Activity of DNA Polymerase θ (POLQ) Is an Intrinsic Property of the Pol Domain and Depends on Unique Sequence Inserts. *Journal of Molecular Biology*, 405(3), 642-652.

[21] Reverter, D., & Lima, C. D. (2004). A Basis for SUMO Protease Specificity Provided by Analysis of Human Senp2 and a Senp2-SUMO Complex. *Structure*, 12(8), 1519-1531.

[22] Solé, A., Delagoutte, E., Ciudad, C. J., Noé, V., & Alberti, P. (2017). Polypurine reverse-Hoogsteen (PPRH) oligonucleotides can form triplexes with their target sequences even under conditions where they fold into G-quadruplexes. *Scientific Reports*, 7, 39898.